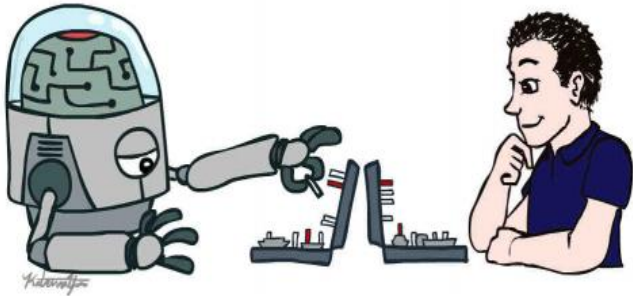
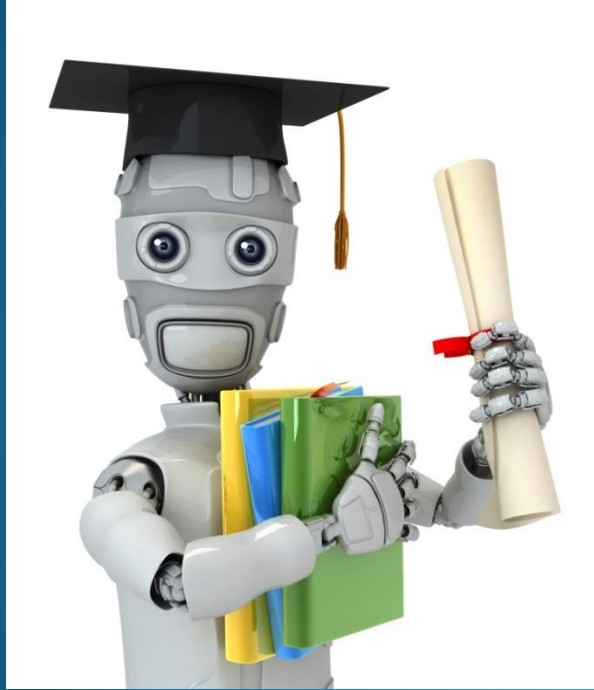


Artificial Intelligence

Part I: Machine learning





Machine Learning

What is machine learning?

Machine Learning

- Grew out of work in AI
- New capability for computers

Examples:

- Database mining
 - Large datasets from growth of automation/web.
 - E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.
 - E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.
- Self-customizing programs
 - E.g., Amazon, Netflix product recommendations
- Understanding human learning (brain, real AI).

Machine Learning definition

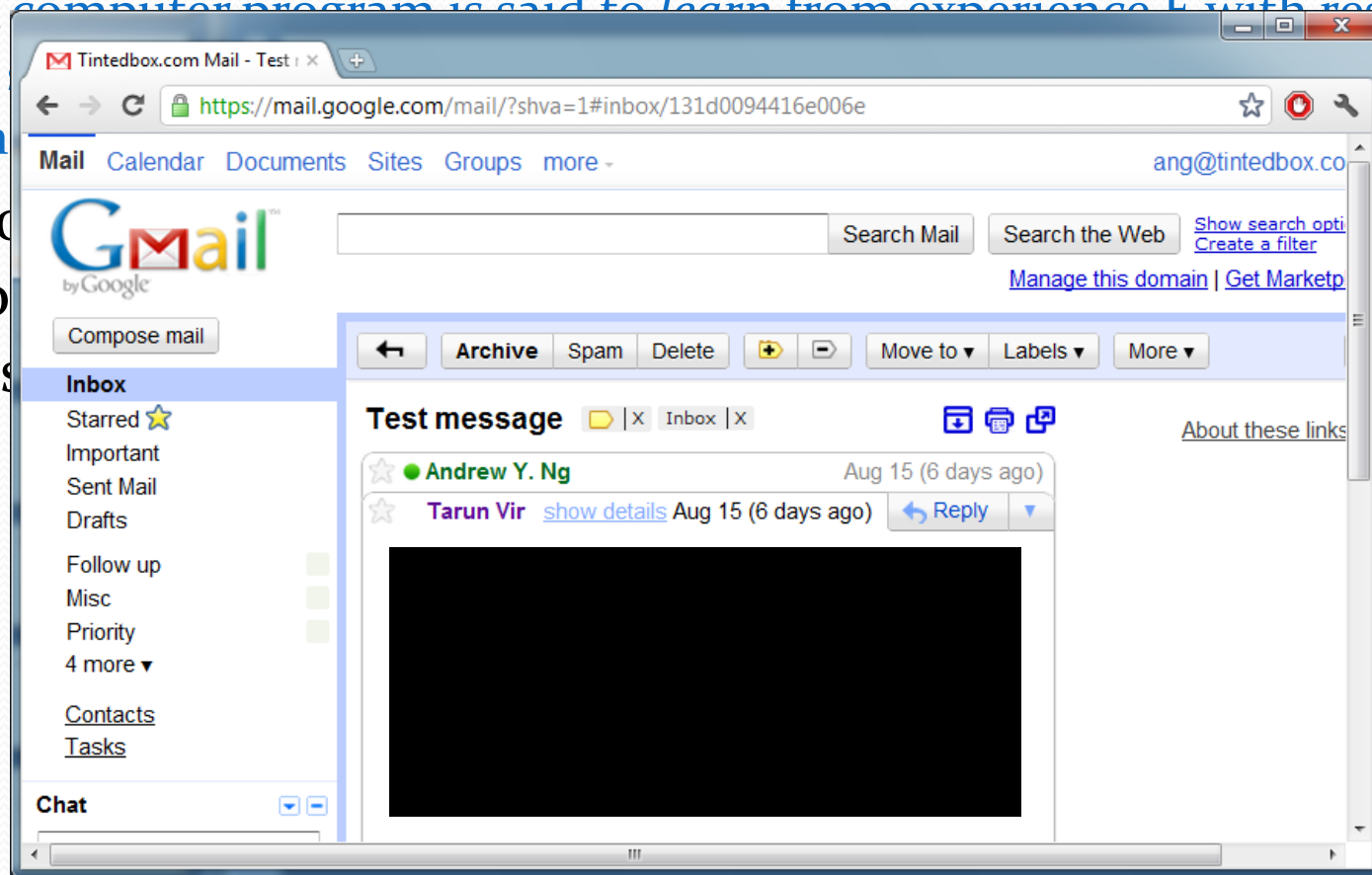
- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
- Tom Mitchell (1998) Well-posed Learning Problem: A computer program is said to *learn* from experience E with respect to some task T and some performance measure P , if its performance on T , as measured by P , improves with experience E .

“A computer program is said to learn from experience E with respect to some task T if its performance on T , as measured by some performance measure P , improves with experience.”

Support
do not
filter

to or
better

spam.

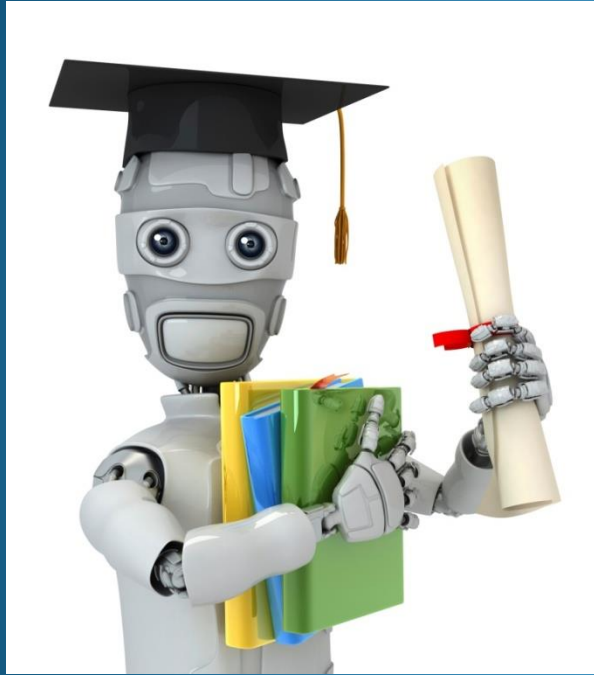




Machine learning algorithms:

- Supervised learning
- Unsupervised learning

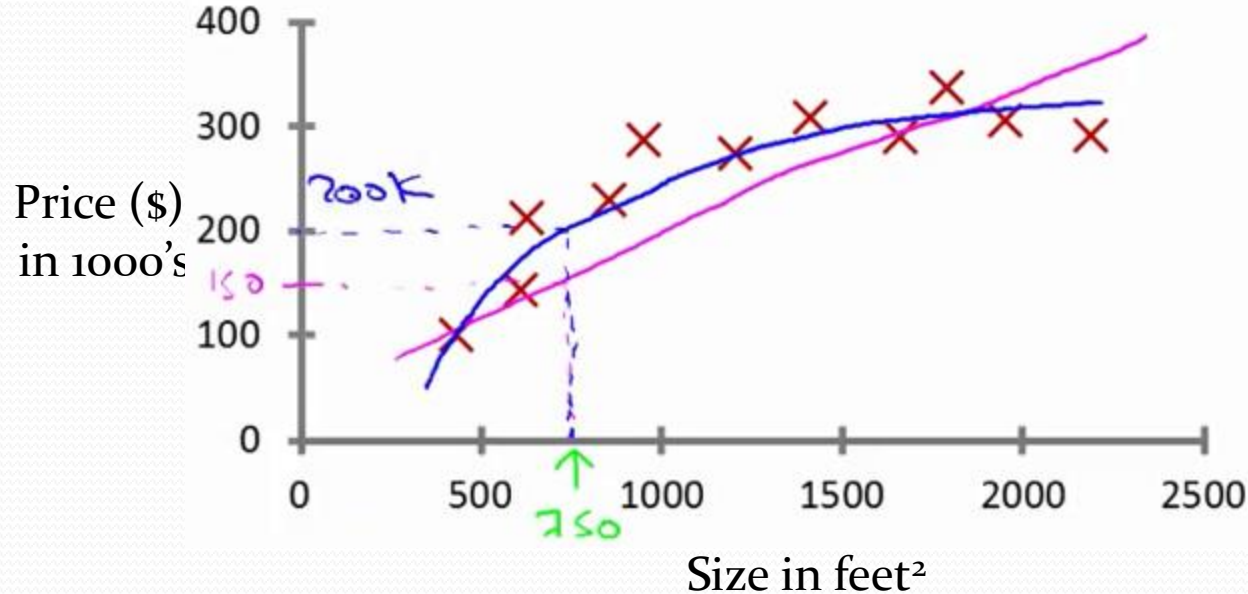
Others: Reinforcement learning, recommender systems.



Machine Learning

Supervised Learning

Housing price prediction.

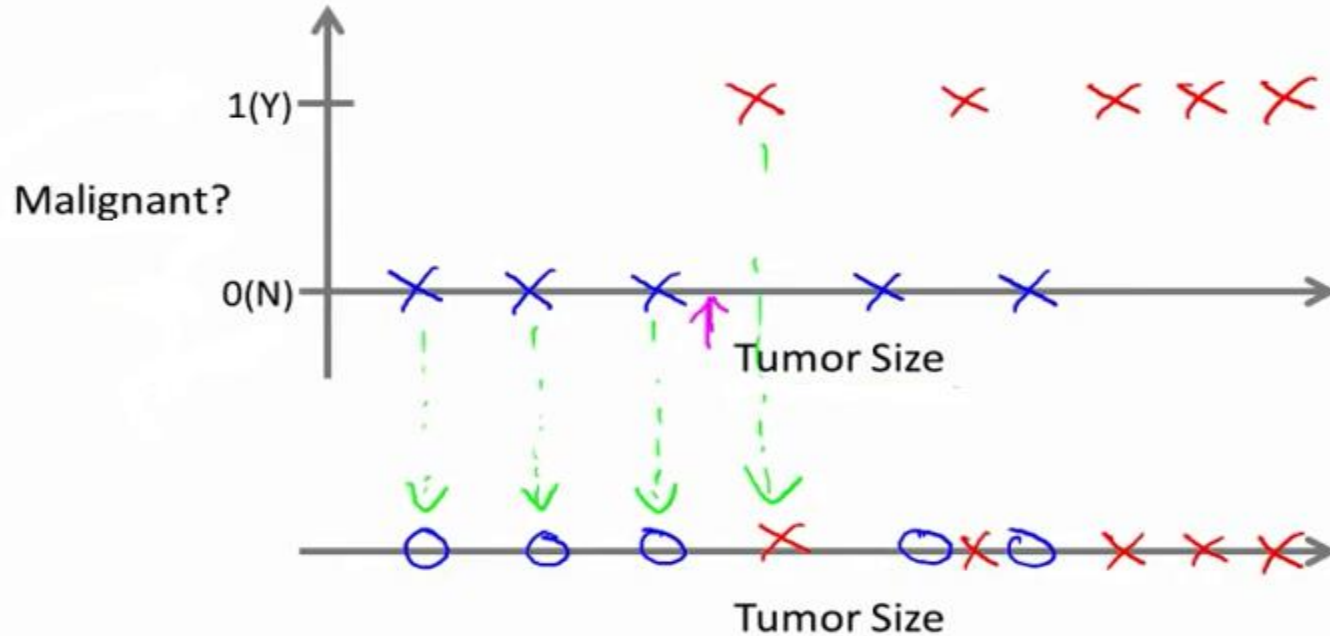


Supervised Learning
“right answers” given

Dr Sherin ElGokhy

Regression: Predict
continuous valued output
(price)

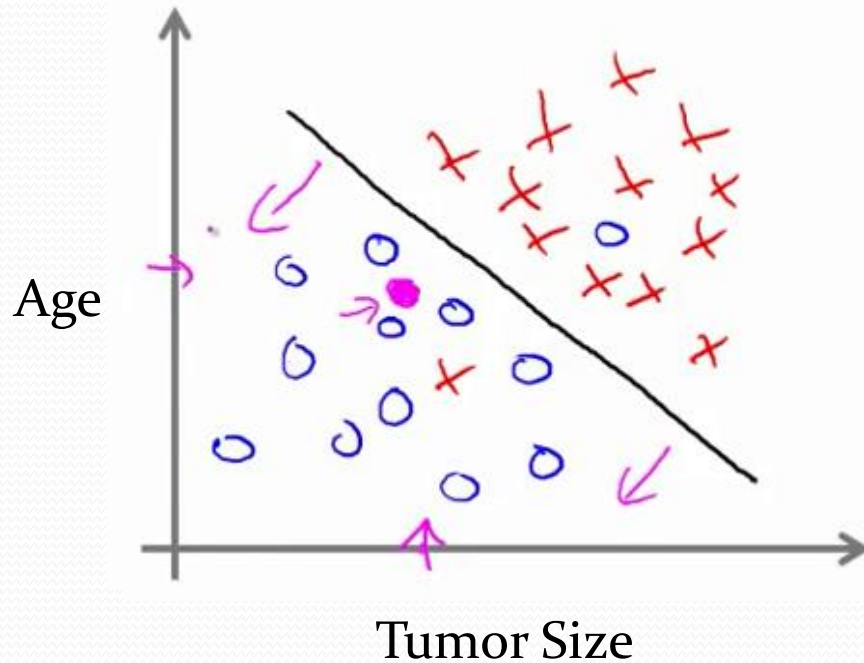
Breast cancer (malignant, benign)



Classification

Discrete
valued output
(0 or 1)

only one feature or one attribute, mainly, the tumor size is used in order to predict whether the tumor is malignant or benign



Clump Thickness
Uniformity of Cell Size
Uniformity of Cell Shape

You're running a company, and you want to develop learning algorithms to address each of two problems.

Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

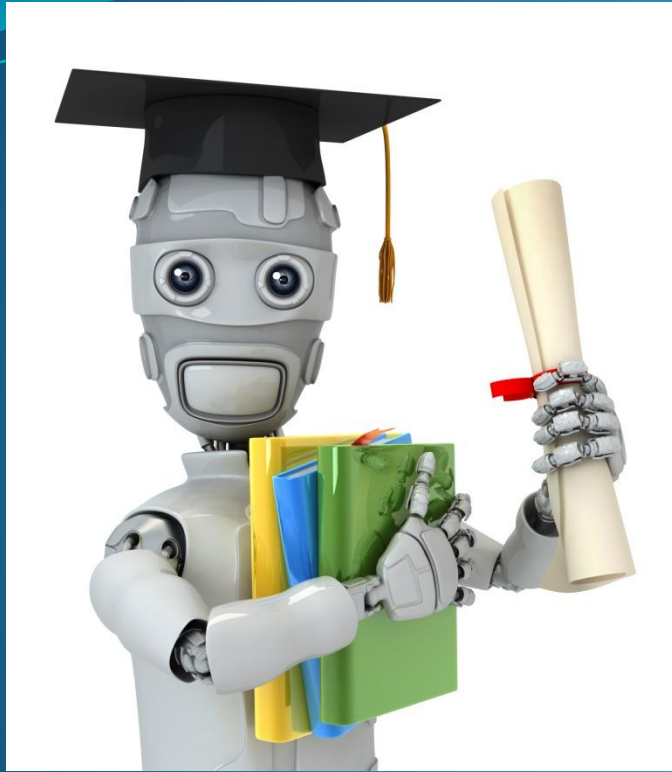
Should you treat these as classification or as regression problems?

- Treat both as classification problems.

- Treat problem 1 as a classification problem, problem 2 as a regression problem.

- Treat problem 1 as a regression problem, problem 2 as a classification problem.

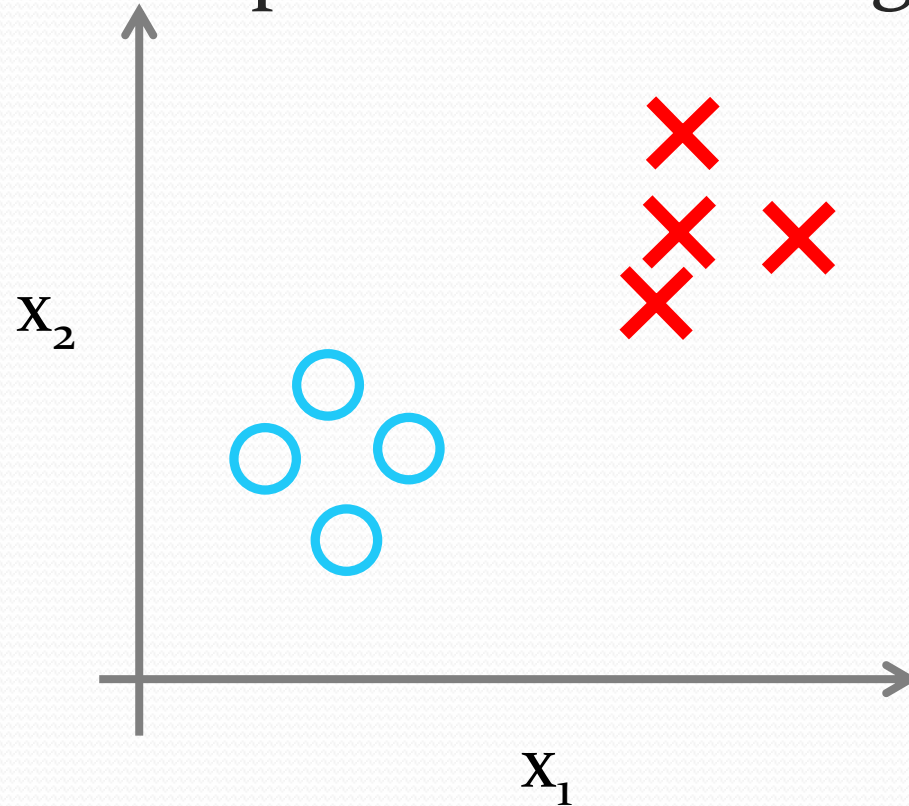
- Treat both as regression problems.



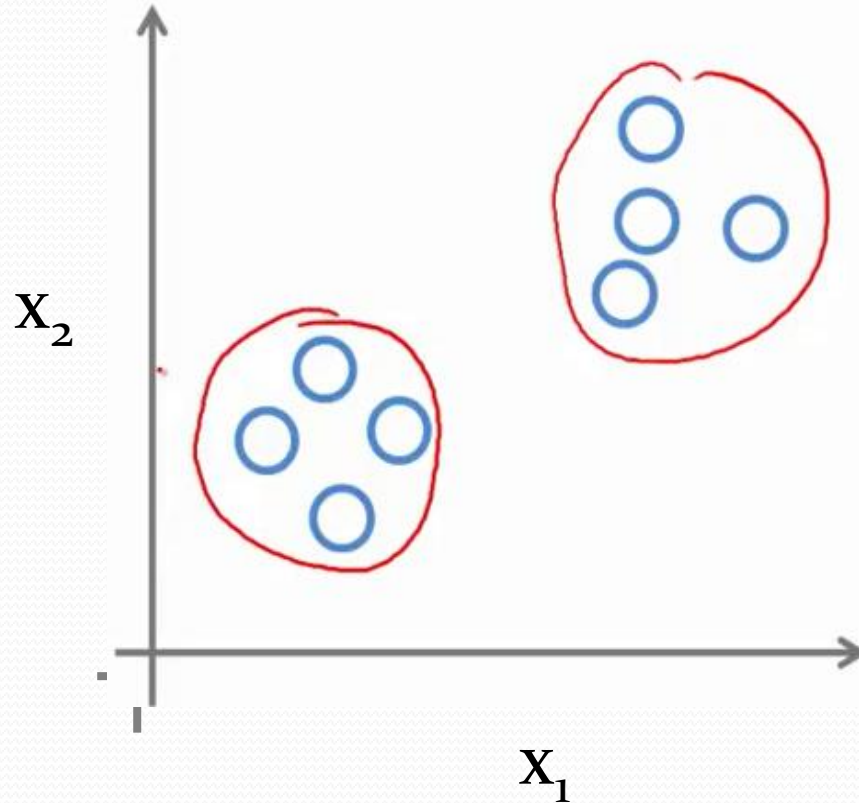
Machine Learning

Unsupervised Learning

Supervised Learning



Unsupervised Learning



Google News
news.google.com

Web Images Videos Maps News Shopping Gmail more ▼

andrewyantakng@gmail.com | Web History | Settings ▼ | Sign out

Google news Search News Search the Web

Advanced news search

U.S. edition ▼ Add a section »

Top Stories

White House official denies Tea Party-focused ad campaign
CNN International - Ed Henry - 1 hour ago
Democratic sources say the White House is not considering an ad campaign tying Republicans to the Tea Party. Washington (CNN) -- A top White House official sharply denied a report that claims President Obama's political advisers are weighing a national ...
Tea Party is misplacing the blame, former President Bill Clinton claims
New York Daily News
GOP tea party backer defends Christine O'Donnell The Associated Press
Atlanta Journal Constitution - Politics Daily - MyFox Washington DC - Salon
all 726 news articles »

US Stocks Climb After Recession Called Over, Homebuilders Gain
MarketWatch - Kristina Peterson - 16 minutes ago
NEW YORK (MarketWatch) -- US stocks climbed Monday, gaining speed after a key nonprofit organization officially called the recession over, giving investors a boost of confidence in the gradual economic recovery.
Longest recession since 1930s ended in June 2009, group says
Los Angeles Times
Downturn Was Longest in Decades, Panel Confirms New York Times
Wall Street Journal - AFP - CNN - USA Today
all 276 news articles »

BP Oil Well, Site of National Catastrophe, Dies at One
Vanity Fair - Juli Weiner - 22 minutes ago
The BP oil well, site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.
Video: Blown-out BP Well Finally Killed in Gulf The Associated Press
Weiss Doubts BP Would End Operations in Gulf of Mexico: Video Bloomberg
CNN International - Wall Street Journal (blog) - The Guardian - New York Times
all 2,292 news articles »

Recent

Recession officially ended in June 2009
CNNMoney - Chris Isidore - 39 minutes ago

Hurricane Igor lashes Bermuda
USA Today - Gerry Broome - 5 minutes ago

'Explain what you want from us.' reads front-page editorial
msnbc.com - Olivia Torres - 10 minutes ago

Crisis response: Pakistan floods

San Francisco Bay Area - Edit

Clorox »
Bay Biz Buzz: Clorox close to selling STP, Armor All
San Jose Mercury News - 48 minutes ago - all 24 articles »

Google's official beekeeper keeps the company buzzing with excitement
San Jose Mercury News - Bruce Newman - 1 hour ago

Jon Sylvia »
Martinez man still unconscious as police investigate weekend shooting
San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

Spotlight

Sarkozy rages at EU 'humiliation'
Financial Times - Peggy Hollinger - Sep 16, 2010

Starred

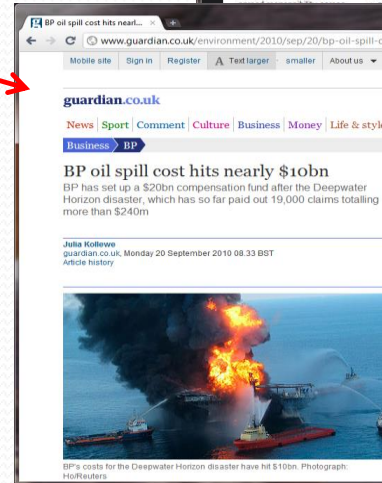
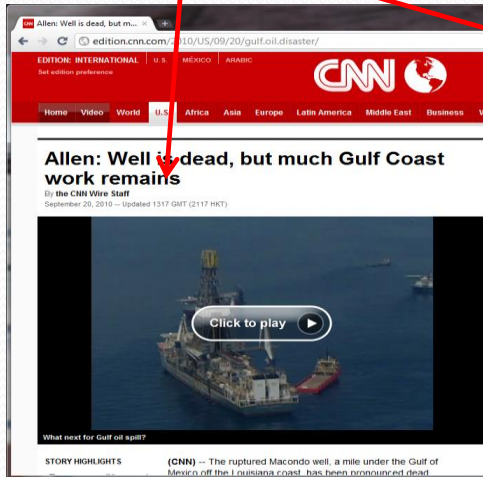
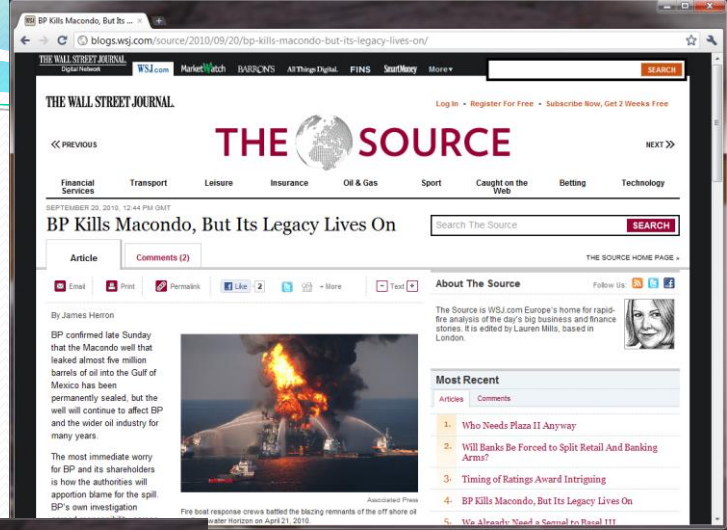
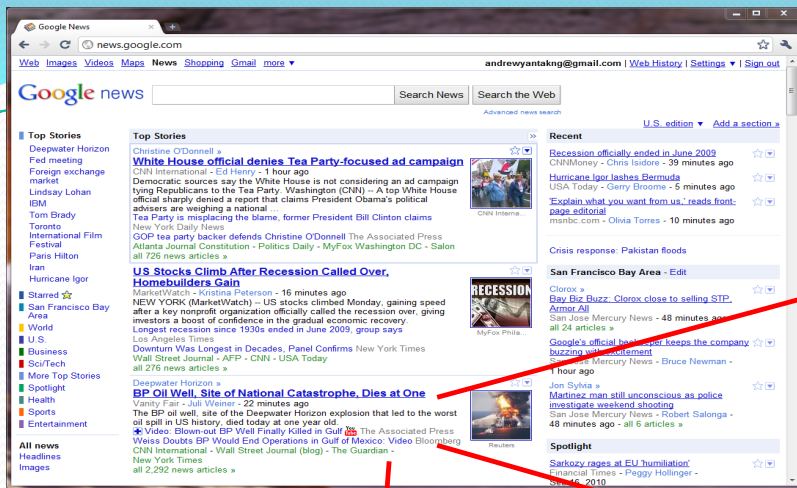
San Francisco Bay Area
World
U.S.
Business
Sci/Tech
More Top Stories
Spotlight
Health
Sports
Entertainment


All news

Headlines
Images

Deepwater Horizon
Fed meeting
Foreign exchange market
Lindsay Lohan
IBM
Tom Brady
Toronto
International Film Festival
Paris Hilton
Iran
Hurricane Igor

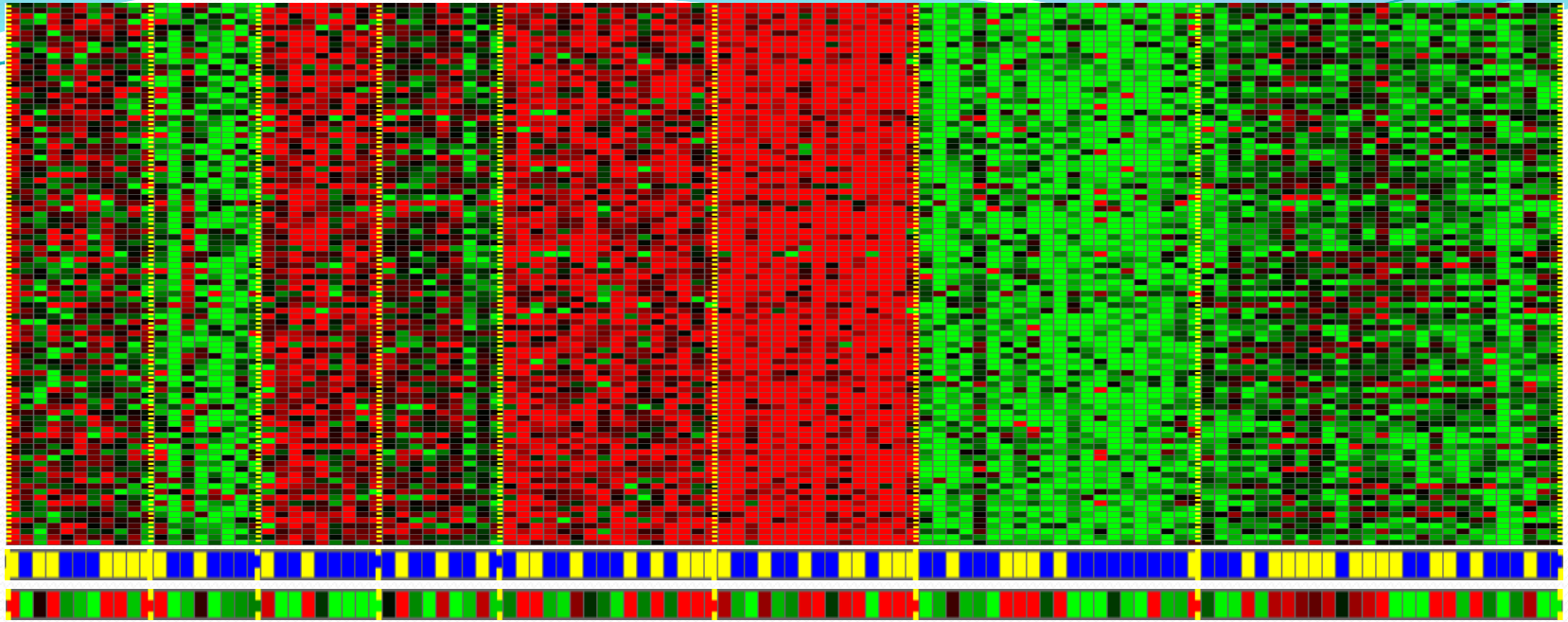
Christine O'Donnell »
CNN Interna...
MyFox Phila...
Reuters





So what Google News has done is look for tens of thousands of news stories and automatically cluster them together.
So, the news stories that are all about the same topic get displayed together.

Genes



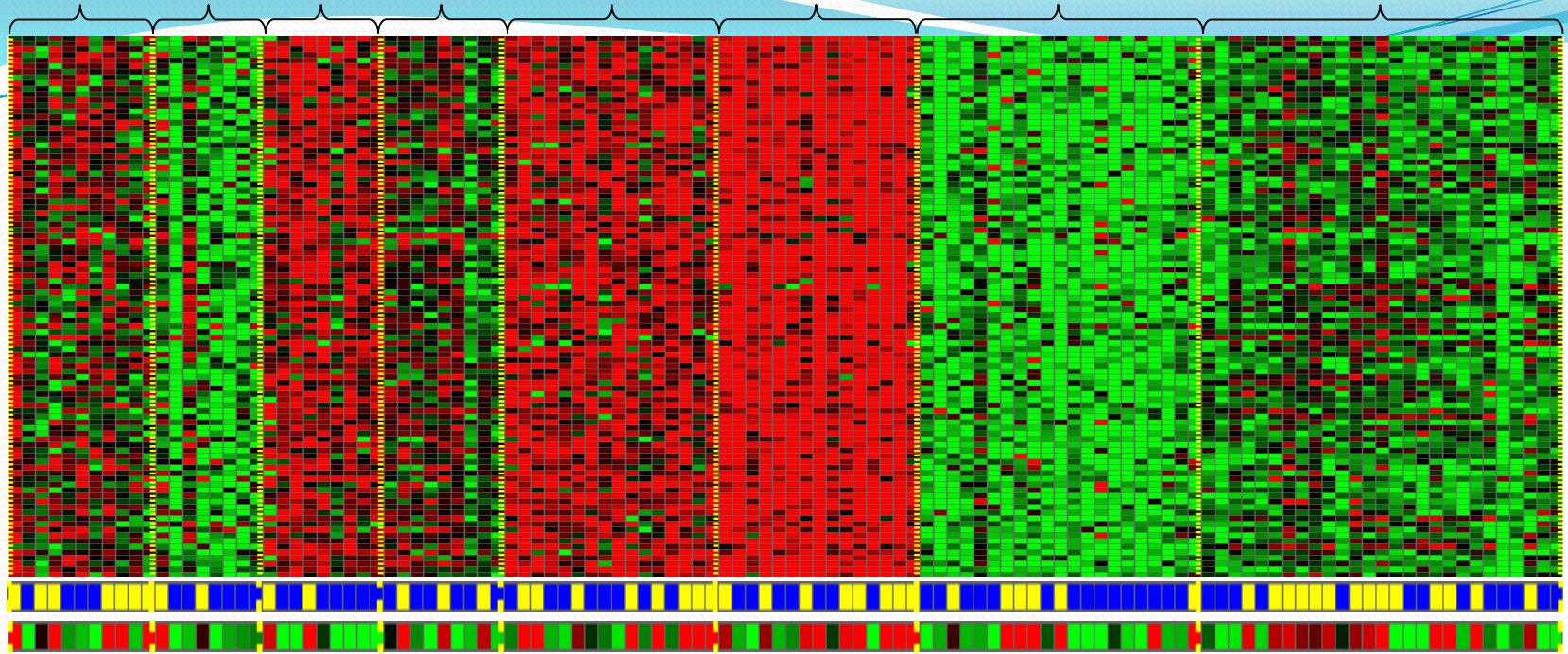
Individuals

Here's an example of DNA microarray data.

The idea is put a group of different individuals and for each of them, you measure how much they do or do not have a certain gene.

Technically you measure how much certain genes are expressed

Genes

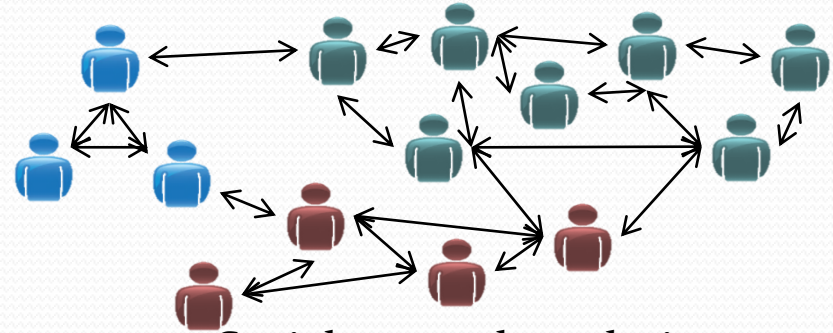


Individuals

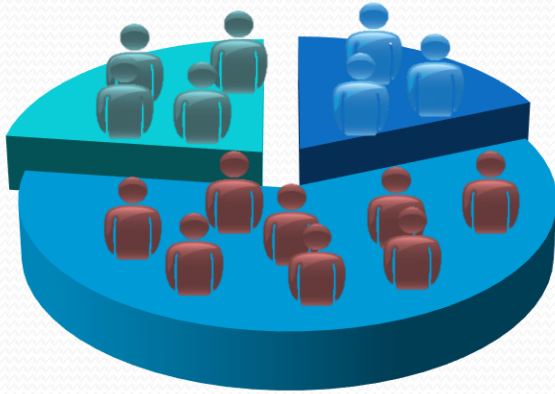
Run a clustering algorithm to group individuals into different categories.
So this is Unsupervised Learning because we're not telling the algorithm in advance that these are type 1 people, those are type 2 persons, those are type 3 persons and so.



Organize computing clusters



Social network analysis



Market segmentation

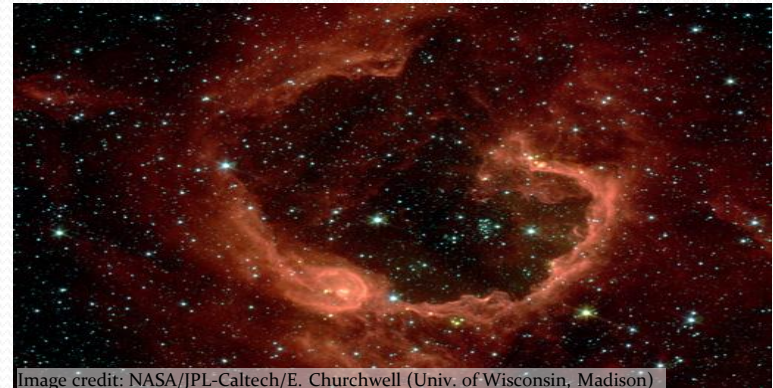


Image credit: NASA/JPL-Caltech/E. Churchwell (Univ. of Wisconsin, Madison)

Astronomical data analysis

Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

Given email labeled as spam/not spam, learn a spam filter.

Given a set of news articles found on the web, group them into set of articles about the same story.

Given a database of customer data, automatically discover market segments and group customers into different market segments.

Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.



Thanks